Michele Scagliarini, Raffaele Vermiglio*

Computing Multivariate Process Capability Indices With Microsoft Excel

*TIM-Telecom Italia Mobile

Quaderni di Dipartimento

Serie Ricerche 2007, n. 1

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Dipartimento di Scienze Statistiche "Paolo Fortunati"

Abstract

In manufacturing industry there is growing interest in measures of process capability under multivariate setting. While there are many statistical packages to assess univariate capability, a current problem with the multivariate measures of capability is the shortage of user friendly software. In this paper a Visual Basic program has been developed to realize an Excel spreadsheet that may be used to compute two multivariate measures of capability. Our aim is to provide a useful tool for practitioners dealing with multivariate capability assessment problems. The features of the program include easy data entry and clear report format.

Keywords: Multivariate Process Capability Indices, Statistical Quality Control, Visual Basic, Excel

1. Introduction

Process capability indices have been widely used in the manufacturing industry providing numerical measures on process performances. Juran *et al.* (1974) first introduced the idea of capability indices (the original name was "capability ratios"). The first indices were univariate, measured the process capability with regard to a single quality measure and focused on the percentage of non-conforming Kotz and Lovelace (1998).

In recent years multivariate capability indices were developed as a natural extension to the univariate concept. Multivariate capability indices appeared in the literature during the early 1990s. Most of them assumed multivariate normal data, a stable process, and were generalizations of their univariate counterparts.

Wang *et al.* (2000) compared three multivariate indices: the multivariate capability vector, Shahriari *et al.* (1995); the multivariate capability index MC_{pm}, Taam *et al.* (1993); the multivariate capability index for process potential MC_p, Chen (1994). While there are many statistical packages for running univariate capability analysis, a current problem with the multivariate measures of capability is the shortage of user friendly software. Recently, an interesting contribution is the work by Phnadnis *et al.* (2005). The authors proposed a Visual Basic program to perform bivariate capability analysis using the MC_{pm} index with Excel.

In this work we have developed a set of Visual Basic macros that may be used to compute multivariate measures of capability using Excel. Since Excel is often used by engineers, or generally by non statisticians, our purpose is to provide a user-friendly tool to help practitioners in performing multivariate capability analysis. We examined two different multivariate capability measures: the multivariate capability vector, Shahriari *et al.* (1995) and the multivariate capability index MC_{pm}, Taam *et al.* (1993). In Section 2 we describe the two multivariate capability measures. The software description is reported in Section 3. Section 4 contains the concluding remarks.

2. Two multivariate capability indices

We assume that a process can be described by a v-dimensional vector of measurements $\mathbf{x}$ and we further assume that the joint probability distribution of the v quality characteristics is the multivariate normal distribution

$$\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{1}$$

2.1. The multivariate capability vector

The multivariate capability vector was proposed by Shahriari *et al.* (1995) based on the original work of Hubele *et al.* (1991). The vector consists of three components and appears as

$$\left(C_{pM}, PV, LI\right)$$

The first component of the vector, C_{pM}, is a ratio of areas or volumes, analogous to the ratio of lengths of the univariate C_p index. The numerator is the area (two-dimensional case) or the volume (three or more dimensions) defined by the engineering tolerance region, while the denominator is the area or volume of a "modified process region", defined as the smallest region similar in shape to the engineering tolerance region, circumscribed about a specified probability contour:

$$C_{pM} = \left[\frac{\text{Volume of engineering tolerance region}}{\text{Volume of modified process region}}\right]^{1/v} \tag{2}$$

where v is the number of characteristics of the process.

The volume of the engineering tolerance region is

$$\prod_{i=1}^{v}\left(USL_i - LSL_i\right) \tag{3}$$

where USL_i and LSL_i are the upper and lower limits respectively, relative to the characteristic i (i=1,2,...,v).

To compute the volume of the modified process region it is worth reminding that under the hypothesis of multivariate normality the statistic

$$(\mathbf{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma} (\mathbf{x}-\boldsymbol{\mu}) = g(\mathbf{x}) \tag{4}$$

follows a χ^2 distribution with v degree of freedom. Therefore, the borders of the process region UPL_i, the upper process limit, and LPL_i, the lower process limit (i=1,2,...,v) are determined by solving the systems of equations of first derivative, with respect to each x_i of the quadratic form

$$(\mathbf{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma} (\mathbf{x}-\boldsymbol{\mu}) = \chi^2_{(v,\alpha)} \tag{5}$$

where $\chi^2_{(v,\alpha)}$ is the upper 100(α) of a χ^2 distribution with v degrees of freedom associated with the probability contour. Usually, in analogy with the "6σ" in the denominator of the univariate indices, α=0.0027.

The solutions (two for each dimension i) of equation (5) are Wang *et al.* (2000):

$$UPL_i = \mu_i + \sqrt{\frac{\chi^2_{(v,\alpha)} \det\left(\boldsymbol{\Sigma}_i^{-1}\right)}{\det\left(\boldsymbol{\Sigma}^{-1}\right)}} \tag{6}$$

$$LPL_i = \mu_i - \sqrt{\frac{\chi^2_{(v,\alpha)} \det\left(\boldsymbol{\Sigma}_i^{-1}\right)}{\det\left(\boldsymbol{\Sigma}^{-1}\right)}} \tag{7}$$

where i=1,2,...,v and $\det\left(\boldsymbol{\Sigma}_i^{-1}\right)$ is the determinant of a matrix obtained from $\boldsymbol{\Sigma}^{-1}$ by deleting the i^{th} row and column. Thus, the volume of the modified process region is

$$\prod_{i=1}^{v}\left(UPL_i - LPL_i\right) \tag{8}$$

In practice $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown and their sample estimates $\overline{\mathbf{x}}$ and $\mathbf{S}$ can be used:

$$\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \tag{9}$$

$$\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n} \left(\mathbf{x}_i - \overline{\mathbf{x}}\right)\left(\mathbf{x}_i - \overline{\mathbf{x}}\right)' \tag{10}$$

Values of C_{pM} higher than one indicate that the modified process region is smaller than the engineering tolerance region, therefore we have high probability that the produced items will be classified as conform.

The second component of the vector is defined as the significance level of a Hotelling T^2 statistic computed under the assumption that the center of the engineering specifications is considered to be the true underlying mean of the process:

$$PV = \Pr\left(T^2 > \frac{v(n-1)}{(n-v)} F_{(v,n-v)}\right) \tag{11}$$

where

$$T^2 = n\left(\overline{\mathbf{x}} - \boldsymbol{\mu}\right)' \mathbf{S}^{-1}\left(\overline{\mathbf{x}} - \boldsymbol{\mu}\right) \tag{12}$$

and $F_{(v,n-v)}$ is the F distribution with v and n-v degrees of freedom.

Values of PV close to zero indicate that the center of the process is far from the engineering target value.

The third component of the vector summarizes a comparison of the location of the modified process region and the tolerance region. It indicates whether any part of the modified process region falls outside the engineering specifications. It has a value of 1 if the entire modified process region is contained within the tolerance region and, otherwise, a value of 0:

$$LI = \begin{cases} 1 & \text{if modified process region is contained within the tolerance region} \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

2.2. The multivariate capability index MC_{pm}

The index MC_{pm} was proposed by Taam *et al.* (1993) and is defined as a ratio of two volumes. The numerator is the volume of the modified tolerance region R_1 and the denominator is the volume of the scaled 99.73 percent process region R_2. Under the multinormality hypothesis we have an elliptical process region, while the modified tolerance region is the largest ellipsoid that is centerd at the target completely within the original tolerance region.

In the general case of v characteristics R_1 is an hyperellipsoid and the volume is given by Kendall (1961)

$$\mathrm{Vol.}(R_1) = \frac{2\prod_{i=1}^{v} a_i}{v} \frac{\pi^{v/2}}{\Gamma\left(\frac{v}{2}\right)} \tag{14}$$

where the a_i (i=1,2,...v) are the lengths of the semi-axes. Then the multivariate capability index is written as

$$MC_{pm} = \frac{\mathrm{Vol.}(R_1)}{\mathrm{Vol.}\left((\mathbf{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}_T^{-1} (\mathbf{x}-\boldsymbol{\mu}) \le K(v) \right)} \tag{15}$$

where $\mathbf{x}$ is the vector ($v\times 1$) of measurements from a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ, $\boldsymbol{\Sigma}_T = E\left[(\mathbf{x}-\mathbf{T})(\mathbf{x}-\mathbf{T})' \right]$ is the mean square error matrix from the process, $\mathbf{T}$ is a vector of target values, and $K(v)$ is a 99.73^{th} percentile of a χ^2 with v degrees of freedom.

The denominator of MC_{pm} can be also expressed as a product of two terms:

$$\mathrm{Vol.}(R_2) = |\boldsymbol{\Sigma}|^{1/2}\left(\pi K(v)\right)^{v/2}\left[\Gamma\left(v/2+1\right)\right]^{-1}\times$$
$$\times\left[1+\left(\boldsymbol{\mu}-\mathbf{T}\right)'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}-\mathbf{T}\right)\right]^{1/2} = \tag{16}$$
$$= \mathrm{Vol.}\left(R_3\right)\times\left[1+\left(\boldsymbol{\mu}-\mathbf{T}\right)'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}-\mathbf{T}\right)\right]^{1/2}$$

where R_3 is the region in which 99.73% of the process values fall within.

Therefore MC_{pm} can be rewritten as:

$$MC_{pm} = \frac{\mathrm{Vol.}(R_1)}{\mathrm{Vol.}(R_3)}\frac{1}{\left[1+\left(\boldsymbol{\mu}-\mathbf{T}\right)'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}-\mathbf{T}\right)\right]^{1/2}} = \frac{C_p}{D} \tag{17}$$

The MC_{pm} index is a function of two components: C_p which represents the process variability relative to the modified tolerance region; D which detects the process deviation from the target. Given a random sample of n measurements, $\mathbf{x}_1$, $\mathbf{x}_2$, ...,$\mathbf{x}_n$ each of dimension v, the estimator for MC_{pm} is given by

$$\hat{M}C_{pm} = \frac{\mathrm{Vol.}(R_1)}{|\mathbf{S}|^{1/2}\left(\pi K\right)^{v/2}\left[\Gamma\left(v/2+1\right)\right]^{-1}}\times$$
$$\frac{1}{\left[1+\frac{n}{n-1}\left(\overline{\mathbf{x}}-\mathbf{T}\right)'\mathbf{S}^{-1}\left(\overline{\mathbf{x}}-\mathbf{T}\right)\right]^{1/2}} = \tag{18}$$
$$= \frac{\hat{C}_p}{\hat{D}}$$

When the process the process mean vector equals the target vector, and the index has the value 1, then 99.73% of the process values lie within the modified tolerance region. The numerator C_p is analogous to the univariate C_p, that is, a value greater than 1 implies that the process has smaller variation than allowed by the specification limits with a certain confidence level; a value less than 1 implies more variation. Also,

$0 < 1/\hat{D} < 1$ measures the closeness between the process mean and the target; a larger $1/\hat{D}$ indicates that the mean is close to target.

3. Software description

The macros are stored in the MultiCap.xls file. The user can directly open this file to perform the capability analysis. To illustrate the software we consider a simulated example. We generate a sample of 100 observation from a multivariate normal process of dimension v=3 with mean vector and covariance matrix given by

$$\boldsymbol{\mu}' = \begin{bmatrix} 40 & 60 & 15 \end{bmatrix}$$

and

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1.100 & 0.483 & 0.308 \\ 0.483 & 0.4 & 0.185 \\ 0.308 & 0.185 & 0.600 \end{bmatrix}$$

respectively.

The target values coincide with the means and the specification limits are reported in Table (1).

Characteristic	LSL_i	LSL_i
1	33	47
2	52	68
3	12	18

Table 1. Specification limits

The user interface is the worksheet "INPUT" (Figure 1) where the main parameters of the analysis can be specified.

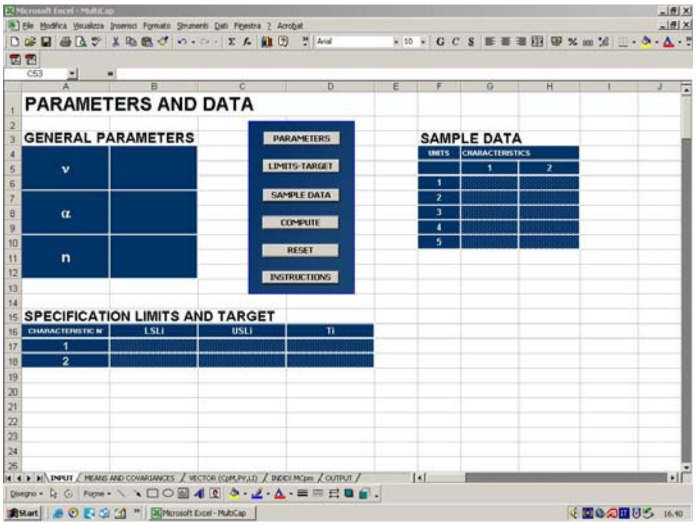

Figure 1. The worksheet INPUT

Clicking the button "PARAMETERS" the form "ANALYSIS PARAMETERS" will appear, as shown in Figure 2, thus can be specified: the number v of quality characteristics *i.e.* the dimensions of the process (for this version the maximum number is 5); the α value to define the size of the tolerance region (usually α =0.0027); the sample size.

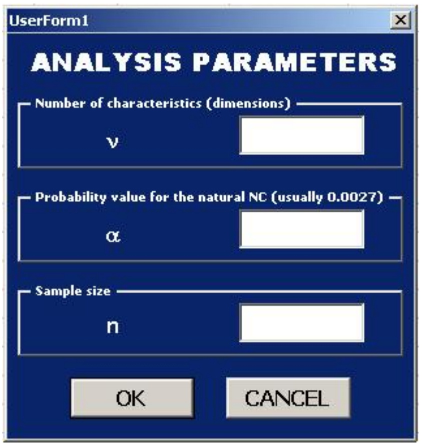

Figure 2. User form "ANALYSIS PARAMETERS"

Therefore, these values will be displayed in the Table "GENERAL PARAMETERS" and the Tables "SPECIFICATIONS LIMITS AND TARGET" and "SAMPLE DATA" will be automatically modified on the basis of the values of *v* and *n*.

Clicking the button "LIMITS-TARGET" the corresponding form (Figure 3) will appear, thus for each quality characteristic can be entered: the specification limits; the target value.

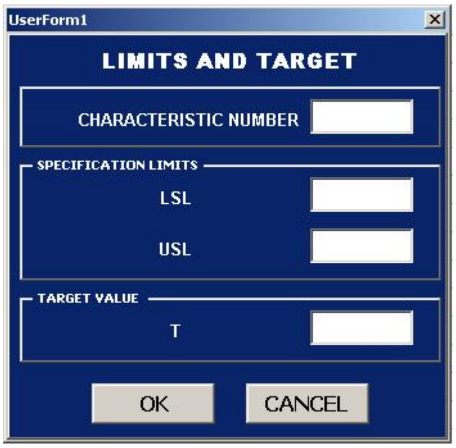

Figure 3. User form "LIMITS AND TARGET"

Clicking the button "SAMPLE DATA" the form "INPUT SAMPLE DATA" (Figure 4) will appear, in this way the sampling observations can be inserted. If the sampling observations are already available, then the data can be directly pasted in the Table "SAMPLE DATA".

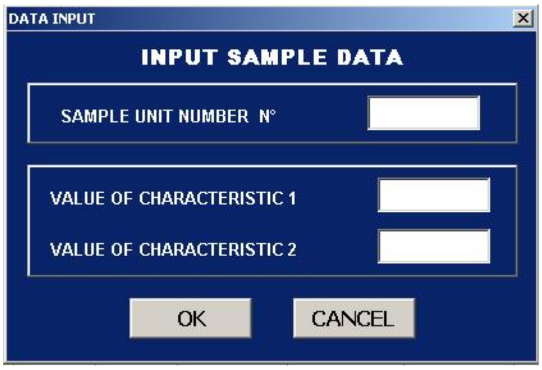

Figure 4. User form "INPUT SAMPLE DATA"

Using the simulated data the worksheet appears as shown in Figure 5.

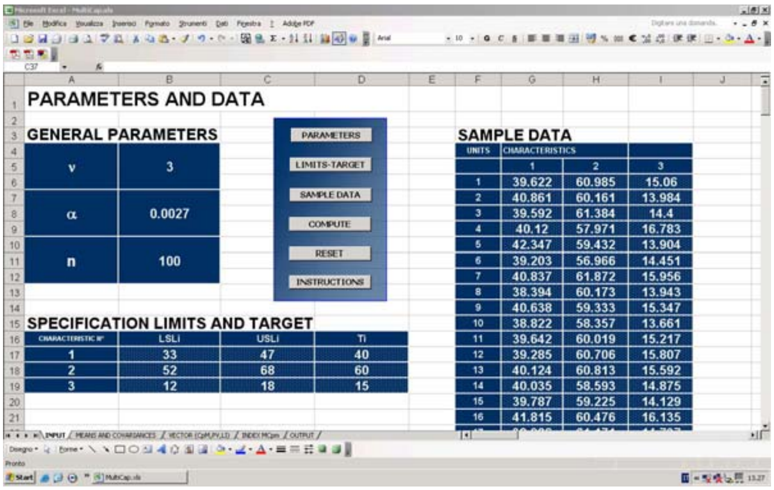

Figure 5. A portion of the worksheet "INPUT" with the example's data

The calculations can be performed clicking the button "COMPUTE". The procedure is splitted in two separate steps. The first step consists in the estimation of the mean vector and covariance matrix. The results of the computations are displayed in the worksheet "MEANS AND COVARIANCES" as shown in Figure 6.

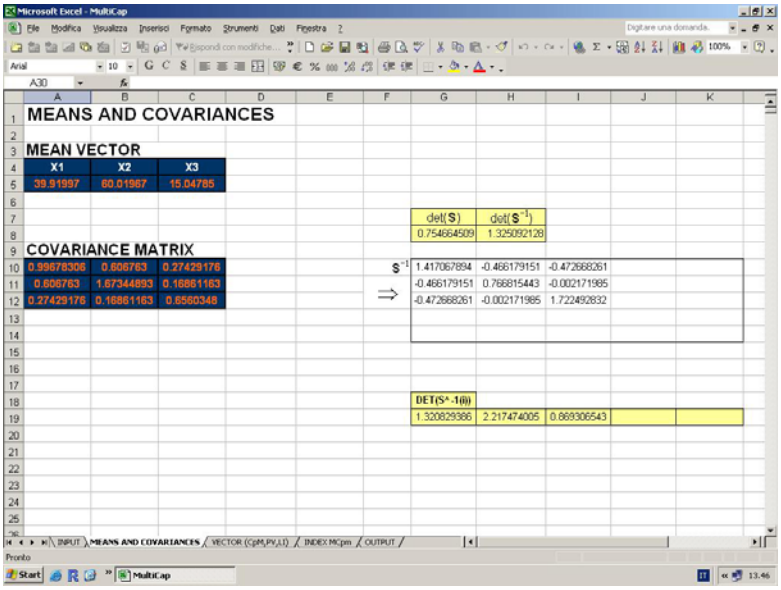

Figure 6. Worksheet MEANS AND COVARIANCES

In the second step the user can choose between the two capability measures (Figure 7). The results will be displayed, together with brief report, in a suitable form (Figures 8 and 9).

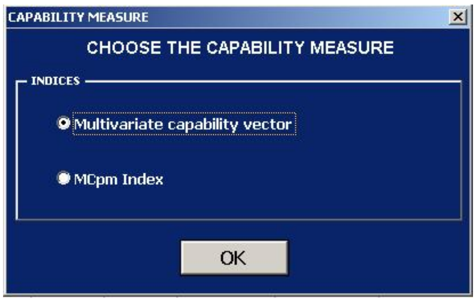

Figure 7. Index choice

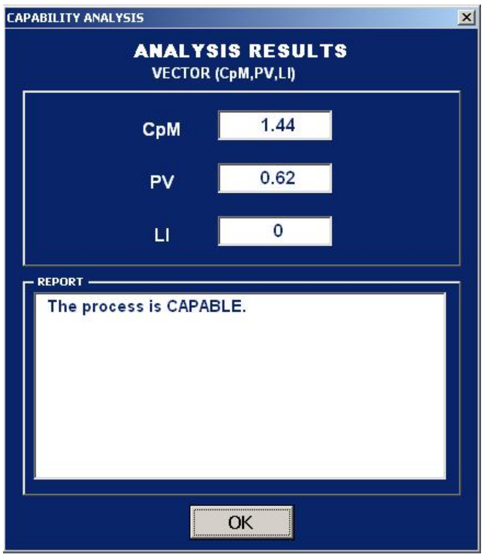

Figure 8. Results and report

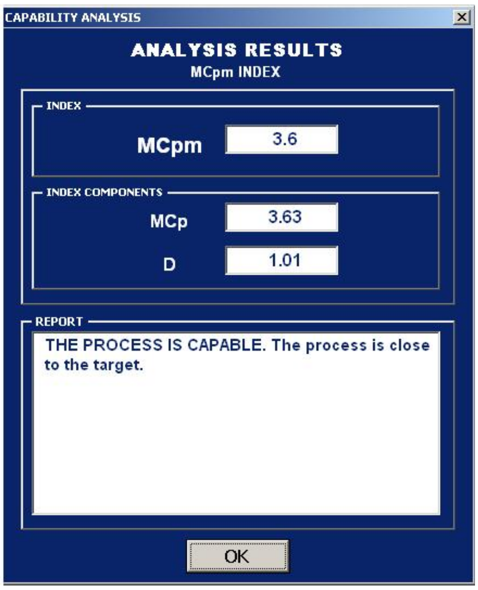

Figure 9. Results and report

In the report box we introduced some comments on the results. The messages reported here are only examples and can be modified if necessary.

In the worksheet "VECTOR" (Figure 10 and Figure 11) are reported the details of the calculations for the multivariate capability vector.

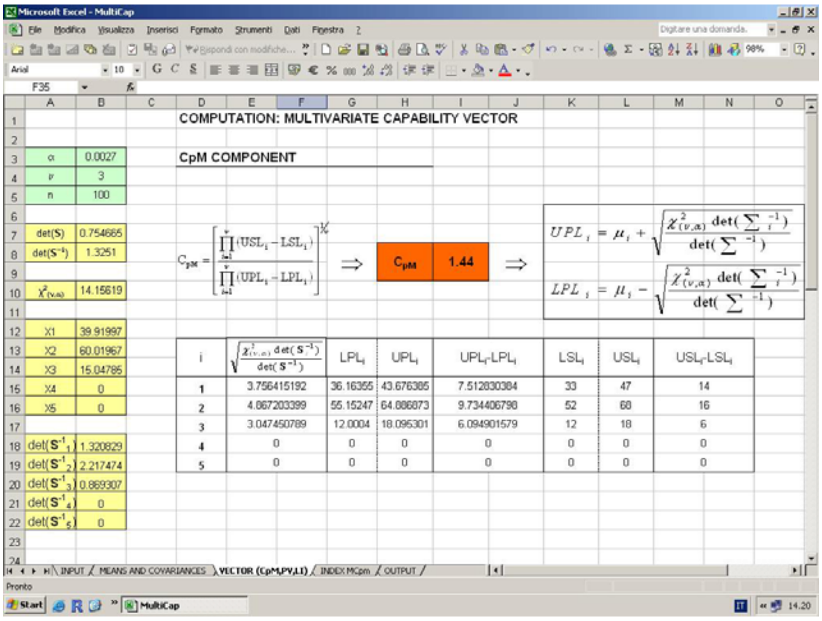

Figure 10. The upper portion of the worksheet "VECTOR"

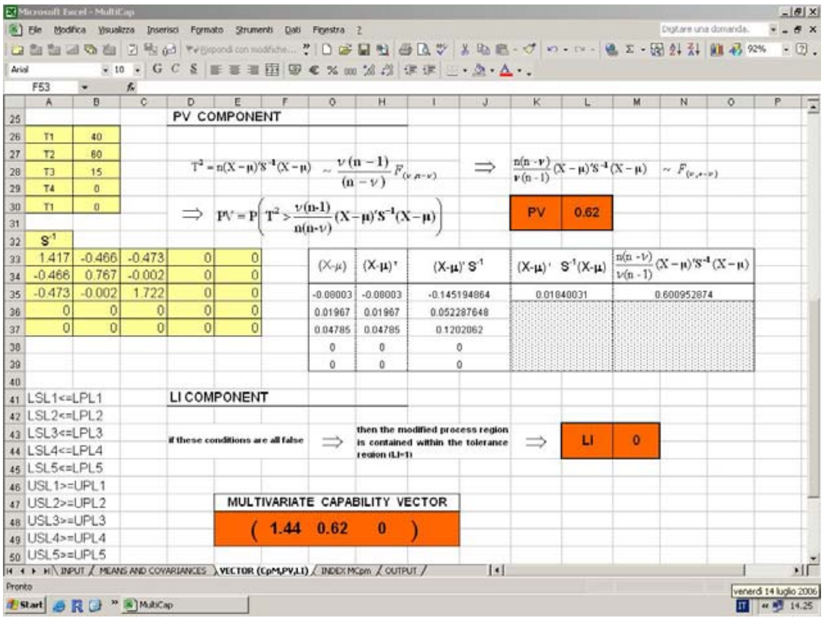

Figure 11. The lower portion of the worksheet "VECTOR"

Similarly, in the worksheet "INDEX MC_{pm}" (Figure 12 and 13) are reported the details of the calculations for the MC_{pm} index.

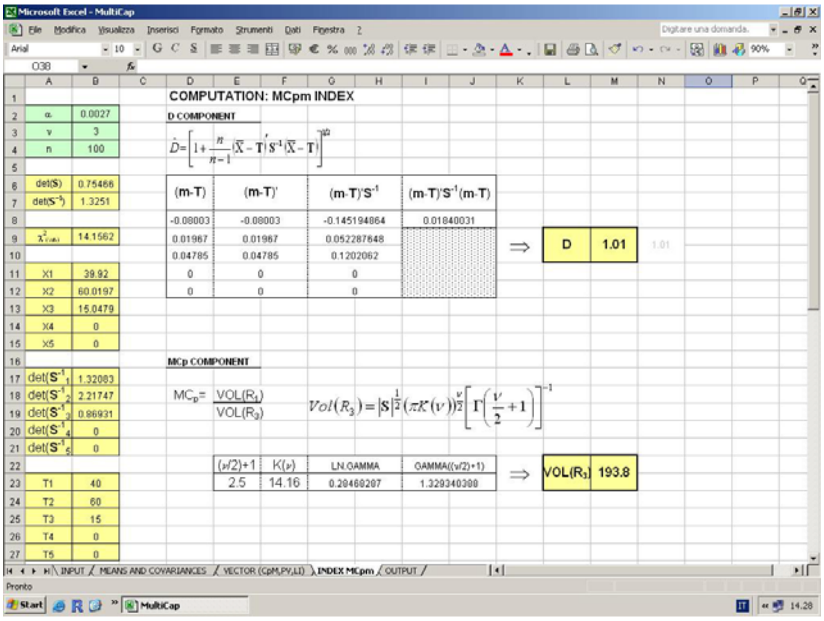

Figure 12. The upper portion of the worksheet "INDEX MC_{pm}"

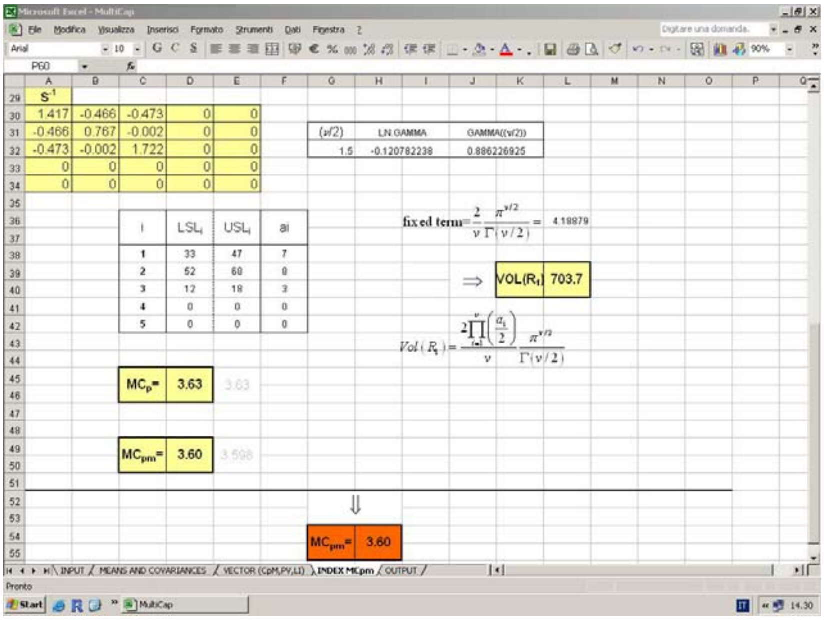

Figure 13. The lower portion of the worksheet "INDEX MC_{pm}"

Results concerning both capability measures are summarized in the worksheet "OUTPUT" as shown in Figure 14.

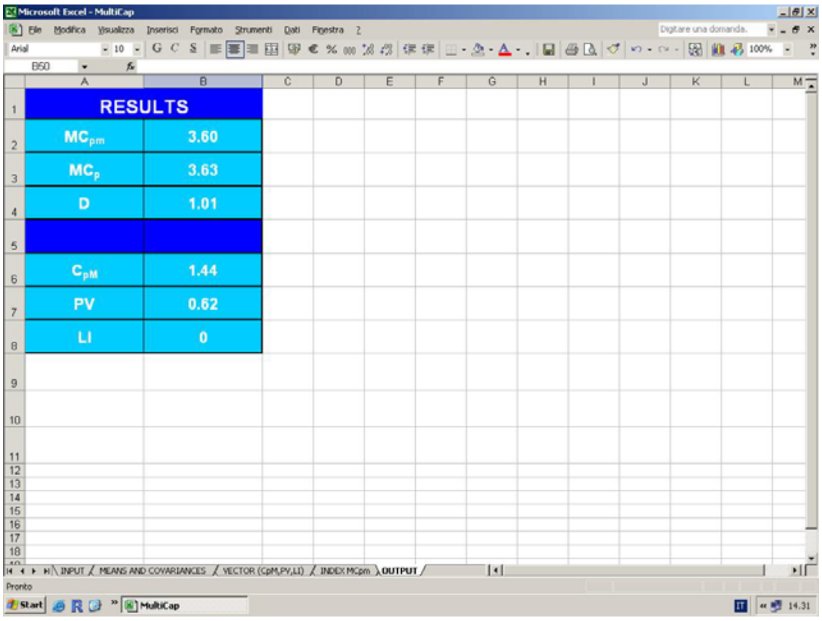

Figure 14. Worksheet "OUTPUT"

3.1. Some details

The Visual Basic code for the calculation management is stored in the Macro CALCOLO1. The inversion of the matrices and the computation of the determinants are performed using the Excel functions MInverse and MDeterm respectively. Moreover, we used the function INV.CHI() to calculate the quantiles of the chi-square distributions in the computations of the process regions and the function DISTRIB.F() to compute the PV component of the multivariate vector. As pointed out by Knusel (1998) and McCullog and Wilson (2002), it is important to take into account the accuracy problems when using these Excel functions. Keeping this caution in mind, in agreement with the purposes of the present work, we retain the degree of precision acceptable.

4. Concluding remarks

In this article we have developed an Excel spreadsheet which can be used to calculate two multivariate capability measures: the multivariate capability vector Shahriari *et al.* (1995) and the multivariate capability index MC_{pm} Taam *et al.* (1993). The proposed software requires no installation, since the user can directly open the .xls file. The spreadsheet interface is easy to use, moreover a set of instructions can be visualized clicking the button "INSTRUCTIONS". Since a problem with the multivariate measures of capability is the shortage of and user-friendly programs we hope that this tool can help practitioners in performing multivariate capability analyses. The software has been validated using several data set. However, the user should understand that there may be undetected bugs and problems and will be grateful for any feedback with relevant comments and suggestions for improvements.

Availability

The file MultiCap.xls file can be obtained freely from the author writing to Michele Scagliarini (scagliarini@stat.unibo.it).

ACKNOWLEDGMENTS

This work has been partly funded by a 2006 MIUR (Italian Ministry for University and Research) grant.

References

H. Chen (1994), *A multivariate process Capability index over a rectangular solid zone,* Statistica Sinica, 4, 749-758.

N.F. Hubele, H. Shahriari, C.S. Cheng (1991), *A bivariate process Capability vector*. In: J.B. Keats, and D.C. Montgomery (Eds.), Statistical Process Control in Manufacturing, Marcel Dekker, New York, 299-310

J.M. Juran, F.M. Gryna, R.S.Jr Bingham (1974), *Quality Control Handbook*. McGraw-Hill, New York.

M. Kendall (1961), *A course in the Geometry of n-dimensions*. Charles Griffin and Company limited, London.

L. Knusel (1998), *On the accuracy of statistical distributions in Microsoft Excel 97*. Computational Statistics and Data Analysis 26, 375-377.

S. Kotz, C. Lovelace, (1998), *Process Capability Indices in Theory and Practice*. Arnold, London.

B.D. McCullough, B. Wilson (2002), *On the accuracy of statistical procedures in Microsoft Excel 2000 and Excel XP*. Computational Statistics and Data Analysis 40, 713-721.

M.A. Phandnis, M.E. Elam, D.J. Fonseca, R.G. Batson, B.M. Adams (2005), *A Prototype DSS for PCI Selection, Calculation, and Interpretation*. Proceedings of the 14th Industrial Engineering Research Conference, URL http://www.bama.ua.edu/~melam/pdss.ht

H. Shahriari, N.F. Hubele, F.P. Lawrence, (1995), *A multivariate process capability vector*. Proceedings of the 4th Industrial Engineering Research Conference, Institute of Industrial Engineers, 304-309.

W. Taam, P. Subbaiah, W.J. Liddy (1993), *A note on multivariate capability indices*. Journal of Applied Statistics, 20(3) 339-351.

F.K. Wang, N.F. Hubele, F.P. Lawrence, J.D. Miskulin, H. Shahriari, (2000), *Comparison of three multivariate process capability indices*. Journal of Quality Technology, 32(3), 263-275.